arXiv:2203.14341v2 [eess.IV] 29 Mar 2022

MFSNet: A Multi Focus Segmentation Network for Skin Lesion Segmentation

Hritam Basak [a], Rohit Kundu [a,*], Ram Sarkar [b]

[a] *Department of Electrical Engineering, Jadavpur University, INDIA*
[b] *Department of Computer Science & Engineering, Jadavpur University, INDIA*

Abstract

Segmentation is essential for medical image analysis to identify and localize diseases, monitor morphological changes, and extract discriminative features for further diagnosis. Skin cancer is one of the most common types of cancer globally, and its early diagnosis is pivotal for the complete elimination of malignant tumors from the body. This research develops an Artificial Intelligence (AI) framework for supervised skin lesion segmentation employing the deep learning approach. The proposed framework, called MFSNet (Multi-Focus Segmentation Network), uses differently scaled feature maps for computing the final segmentation mask using raw input RGB images of skin lesions. In doing so, initially, the images are preprocessed to remove unwanted artifacts and noises. The MFSNet employs the Res2Net backbone, a recently proposed convolutional neural network (CNN), for obtaining deep features used in a Parallel Partial Decoder (PPD) module to get a global map of the segmentation mask. In different stages of the network, convolution features and multi-scale maps are used in two boundary attention (BA) modules and two reverse attention (RA) modules to generate the final segmentation output. MFSNet, when evaluated on three publicly available datasets: PH^2, ISIC 2017, and HAM10000, outperforms state-of-the-art methods, justifying the reliability of the framework. The relevant codes for the proposed approach are accessible at

*Corresponding author
Email address: rohitkunduju@gmail.com (Rohit Kundu)

(a) Raw dermoscopy image (b) Ground truth mask

Figure 1: Example of a skin lesion image and its ground truth mask.

https://github.com/Rohit-Kundu/MFSNet

Keywords: Lesion Segmentation, Deep Learning, Parallel Partial Decoder, Attention Modules, Skin Melanoma

1. Introduction

Melanoma is the most severe and deadly type of skin cancer, causing more than 13 thousand incidences globally. Though less prevalent than its non-malignant counterpart, malignant melanoma is increasing at an alarming rate of 4% per year. Research has shown its correlation with genetic and physical variations. The primary cause of melanoma is long-term exposure to ultraviolet (UV) rays. With the increase in greenhouse gases, the protective ozone layer in the stratosphere is depleting rapidly, causing the harmful solar UV rays to reach the earth's surface. This causes the global incidence of melanoma to rise rapidly. Fortunately, studies like Siegel et al. [40] show that early detection can decrease the chances of fatality by 97%. Surgical treatment of melanoma is often disfiguring and extremely painful, justifying the importance of early detection of the disease. Dermoscopy is a non-invasive test for detecting and diagnosing pigmented skin lesions and malignant melanoma in the early stages. It is often considered the golden standard for melanoma localization. However, manual labeling and reviewing are extremely grueling and cumbersome even for expert clinicians, relying on their perceptions and vision. Therefore, to mitigate the problem, Computer-Aided Diagnosis (CAD) systems have been widely preferred

as a support system to aid clinicians in automated segmentation and analysis of malignant melanoma.

Semantic segmentation refers to the pixel-level classification of the images. Each pixel in an image is classified as part of the object class or background class. This is beneficial for localizing the region of interest (ROI) from the raw images for further analysis and thus is a vital preprocessing step in automated disease diagnosis. Figure 1(a) shows an example of a raw skin-lesion image. Its segmented image, called "ground truth," is shown in Figure 1(b). Here, the image is classified into two classes, namely "lesion" and "background," which led to the generation of a "binary mask" image. The task of semantic segmentation is to generate a segmentation map like Figure 1(b) from raw input image similar to Figure 1(a). To address this, extensive research attempts have been made since the last decade to automate the segmentation of lesions, monitor their growth, and aid physicians in making surgical decisions, thereby increasing the clinical significance.

Segmentation of skin melanoma from non-invasive dermoscopy images relies upon several emerging and traditional methods. Among them, segmentation methods based on artificial intelligence have widely been explored and adopted due to their excellent accuracy, robustness, and reliability. Extensive research has been conducted in the last few years, using neural networks [4], fuzzy logic [21], attention-gated networks [48], or their combinations with traditional image processing methods to improve the segmentation performance. The significant variations in texture, size, shape, the position of lesions, and obscure boundaries in dermoscopy images make it extremely challenging to obtain accurate and prominent tissue-level segmentation maps for developing CAD systems. Pre and post-processing are the other essential aspects and used in most of the current segmentation methods [14] for effectively removing artifacts, enhancing image quality, removing unnecessary noises in images for effective and accurate segmentation of pigmented skin lesions from images. Beuren et al. [8] proposed a series of morphological operations for image enhancement of image resolution and denoising before segmentation. Later Chatterjee et al. [14] proposed the

Fractal Region Texture Analysis (FRTA) method for quantification of texture information integrated with Recursive Feature Elimination (RFE) and several morphological operations as preprocessing before classification of dermoscopic images. Verma et al. [46] showed that median filters and anisotropic diffusion filters can be helpful in not only smoothing the images but also removal of thick hairlines, preserving sufficient lesion edge information. Recently, morphological operations and image inpainting methods have been modified and used in research for dermoscopy image analysis [38]. In the preprocessing step, this research has incorporated the image inpainting method for unwanted hair removal from the input images.

In literature, the skin lesion segmentation methods are broadly classified into the following categories: (a) edge detection and thresholding [31], (b) active contour models [45], and (c) segmentation based on convolutional neural network (CNN) [50, 7], etc. Symmetrical encoder-decoder architecture, also known as U-Net, proposed by Ronneberger et al. [36], is widely used and considered as the golden standard for several image segmentation tasks. It consists of a downsampling path that captures sufficient semantics and context, connected to an expanding path for accurate localization of the ROI. Later Zhou et al. [57] proposed a novel architecture UNet++ by redesigning the series of nested dense skip connections to reduce the semantic gap between the feature representations and the encoder-decoder sub-networks. Their proposed model outperformed the previous U-Net architecture in multiple biomedical image segmentation tasks. Weng et al. [51] proposed another modification in the U-Net backbone by incorporating neural architecture search (NAS), thereby improving the segmentation performance significantly. SegNet [5] is a similar encoder-decoder model, for instance, segmentation, that uses a VGG16 backbone followed by a decoder path integrated with a pixel-wise classification layer. This is a well-known segmentation model for binary or multi-class segmentation problems and has been proven to produce state-of-the-art results in various domains. Yuan et al. [54] proposed a fully convolution-deconvolution network that was able to produce a dice similarity score of 76.5% on the ISIC 2017 dataset. Later Abraham et al.

[1] proposed a novel Focal Tversky Loss function, then integrated with attention U-Net, produced a state-of-the-art result on BUS 2017B and ISIC2018 datasets with average dice scores of 80.4% and 85.6% respectively. Double U-Net [26], another modification of U-Net that used two different upsampling branches instead of one, was used to produce two different segmentation maps, slightly different from one another. The paper reported an average dice score of 89.2% on the ISIC 2017 dataset.

Recently meta-heuristic-based optimization algorithms have been explored for thresholding-based segmentation and image enhancement operations in different applications. Aljanabi et al. [3] proposed an image thresholding method by selecting an optimum threshold level using the artificial bee colony (ABC) algorithm. The algorithm was able to produce segmentation maps with high confidence on several widely known skin datasets. Attention mechanisms are also widely known for boosting the performance of CNN-based models in different computer vision applications. Chattopadhyay et al. [15] proposed a multi-scale attention mechanism which is inspired by the work of [9] for accurate localization and segmentation of objects. The dual attention mechanism was proposed by [19] adaptively integrates the local features with their corresponding global dependencies. Though used in scene segmentation application, it inspired several similar works in the biomedical domain [6]. Generative Adversarial Networks (GAN) have also been instrumental for extensive research for biomedical image segmentation recently [27].

1.1. Overview and Contributions

To address the issues mentioned before, we propose a novel skin lesion segmentation framework, called Multi-Focus Segmentation Network (MFSNet), that produces the final segmentation map by focusing on image information at multiple scales. Taking a clue from the standard clinical practice, we can say that the area and boundary are the two essential aspects to produce the accurate pixel-level segmentation map based on local appearance from a coarse localization of the melanoma region. The proposed model generates a coarse segmen-

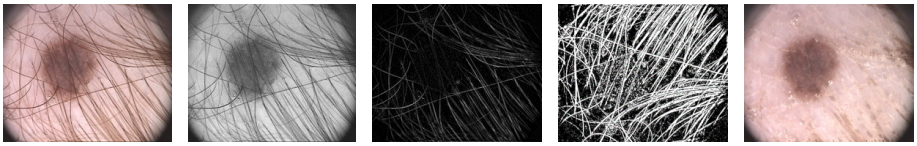
tation map implicitly by aggregating image features at multiple levels, followed by a series of reverse and boundary attention networks by iteratively learning pixel-level information of area and boundary by explicitly using the coarse map and ground truth the global guidance. We have evaluated the performance of the proposed model on three publicly available skin melanoma datasets: The PH^2 dataset by Mendoncca et al. [33], the ISIC 2017 dataset by Codella et al. [17] and the HAM10000 dataset by Tschandl et al. [42]. The proposed model outperforms state-of-the-art models on the same datasets justifying the reliability and robustness of the framework.

The contributions of the present research are as follows:

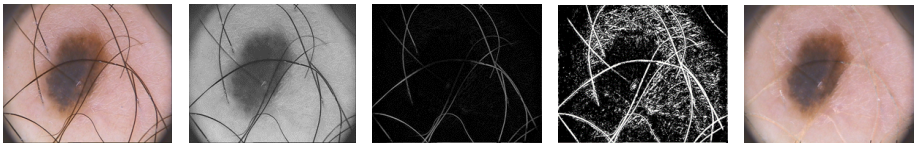
1. The use of differently focused segmentation maps in various stages of the proposed MFSNet helps accurately map both the lesion's coarse structure and its fine edges.
2. Unlike the commonly used segmentation frameworks in literature, the proposed model upsamples the encoded features in subsequent steps of attention modules instead of coarse upsampling applied in U-Net type architectures.
3. We evaluate the proposed MFSNet model on three publicly available datasets: PH^2, ISIC 2017 and HAM10000 datasets, and obtain dice similarity coefficient values of 0.954, 0.987, and 0.906 respectively on the datasets, outperforming state-of-the-art methods on the same datasets.

2. Proposed Method

This section describes the architecture of our proposed MFSNet, which combines the high-level semantics and the low-level edge information by using a series of RA modules, BA block, and a PPD module. We propose a hybrid loss function that integrates the widely used Binary Cross-Entropy (BCE) loss with the Weighted IoU loss functions. The whole segmentation process is followed by image inpainting and a preprocessing step for artifact removal, described in Section 2.1.



(a) Original Image (b) Grayscale Image (c) Blackhat Filtered Image (d) Threshold for Inpainting (e) Final Inpainted Image



(f) Original Image (g) Grayscale Image (h) Blackhat Filtered Image (i) Threshold for Inpainting (j) Final Inpainted Image

Figure 2: Outputs of the image inpainting method used for artefact removal on the PH2 dataset: (a) & (f)- Original images; (b) & (g)- Corresponding grayscale images; (c) & (h)- Blackhat filtered images; (d) & (i)- Thresholding for the inpainting operation; (e) & (j)- final preprocessed (inpainted) images.

2.1. Image Preprocessing

Dermoscopy images vary in terms of size, pixel intensity and may suffer from unwanted artifacts in the form of noises or body hair. These artifacts may lead to abrupt segmentation results in some images and may diminish the overall model performance. Hence, to address these problems, we used standard image preprocessing methods before segmenting the images. All the images have been resized to a shape of 256×256 for faster convolution and resolving excessive memory constraints. Next, we perform image normalization to resolve the uneven image contrast issues. Finally, we introduce the image inpainting method for hair removal.

Following the work of [41], we have used several morphological operations for hair removal from dermoscopy images. First, the input RGB images are converted to grayscale images, followed by blackhat transformation as proposed by [47]. In this regard, we define a structuring element: a cross-shaped two-dimensional array of shape 17×17, i.e., an array whose middle row and column are composed of 1's and all other places contain 0.

Similar to [47], closing is also performed to remove small hollows inside a region while keeping the original region shape and size unaltered. Thus the blackhat transformation results in an output image containing elements darker than the surrounding pixel values, whereas smaller than the structuring element. A suitable threshold value is applied to the obtained output from the blackhat transformation to obtain the hair-like artifacts.

Fast marching method [30] is widely used for segmentation purposes. In this research, we have used this algorithm for image inpainting. We used the thresholded image output from blackhat transformation and the original input image and replaced the artifacts or hair structures with the neighboring pixels. Figure 2 shows the image outputs from different intermediate steps of image artifact removal.

2.2. MFSNet Architecture

Figure 3 shows the architecture of the proposed MFSNet. It consists of the Res2Net as a backbone, which is a recently proposed CNN model [20], for feature extraction combined with a series of RA branches, explained in Section 2.4. Only five initial convolution layers of the network are used for this purpose. The first three layers are used to extract low-level features with high resolution but very little spatial information. The second and third level features, F_2 and F_3, with important edge information, are fed to the BA module to improve melanoma boundary representation. F_2 and F_3 are further used for two different purposes. They are fed to the following two layers of the CNN, whose output is fed to the PPD module to generate the global segmentation map O_S, which is used as the global map for coarse localization of melanoma segmentation. Secondly, they are fed to the RA branches, along with O_S, to be used as the global guidance for the entire learning process of the network. The subsequent two layers of features $F_i; i = 4, 5$ from the successive two consecutive layers of the CNN are fed to the corresponding RA module to produce output $O_R(F_i)$,which is concatenated with the upsampled $O_R(F_{i+1})$ from the next branch, thereby ensuring the multi-level feature representation. This results in the output O_i from each branch, which is

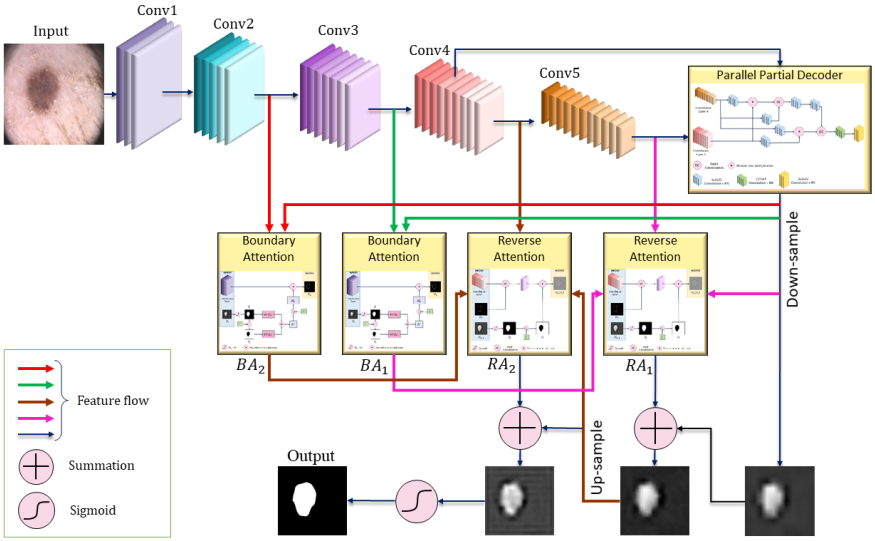


Figure 3: Overall structure of the proposed MFSNet model for the segmentation of skin lesions. The inputs to the boundary attention and reverse attention blocks have been shown in different colored arrows. The inputs of BA_1, BA_2, RA_1, and RA_2 are marked using green, red, pink and brown arrows respectively.

supervised with the ground truth G through a loss function (described in Section 2.7). Thus, by using the parallel RA and the residual connections between the segmentation of multiple scales and the ground truth, the errors can be removed by "larger-scale adaptability" [16]. Finally, the output O_2 is passed through the sigmoid activation function to produce the final segmentation map S.

In general, the error between the input and output of the RA unit is minor (zero in the extreme case), thus making the learning comparatively easy with very few parameters. Hence, the network can be very effective in region segmentation with fewer parameters. As the learning procedure of the network is focused on generating multiple levels of outputs from multiple branches, the network is named as MFSNet.

2.3. Workflow of the MFSNet

As shown in Figure 3, the proposed model consists of a series of convolution operations, RA branches, and BA modules. For the convenience of the readers, we have described below the flow of information from the input image through the layers and branches to produce the segmentation output finally.

1. The input image is initially passed through a series of convolution layers for feature extraction using the Res2Net backbone, where downsampling is performed. Among those, only the features of the second and third Convolution layers are considered useful for edge guidance of the learning process because the low-level features preserve sufficient boundary information [55]. Hence they are used for the BA module that explicitly learns the boundary information. Upsampling is performed in the PPD module.

2. The BA module simultaneously takes input from the global segmentation map (output from PPD) and the shallow features from the convolution layers. By performing a series of distance transformations and other mathematical operations, an enhanced boundary map is obtained, further used by the RA modules. The detailed algorithm and workflow of BA are described later in Section 2.5.
3. The RA module takes the features from the corresponding convolution layer, BA module, and the upsampled segmentation map from the next layer. The RA module uses two separate input branches to learn features to produce segmentation masks associated with two different classes - foreground and background. Thus the RA module generates a per-class mask to amplify the reverse-class response in the regions that contain high-level semantic information shared between two adjacent classes. Finally, the prediction of these two branches is fused to generate the segmentation output from the RA branch. The detailed workflow of the RA module is explained in Section 2.4.

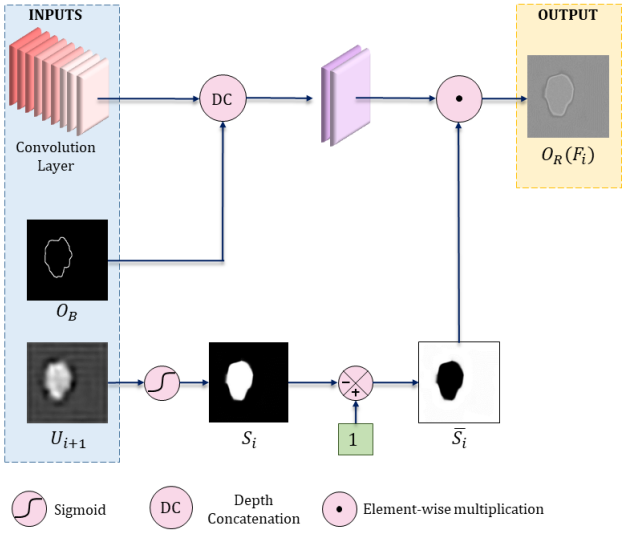


Figure 4: Architecture of the RA module used in the proposed MFSNet model.
O_B: Output from the BA module; U_{i+1}: Upsampled output from the next layer; $O_R(F_i)$: Output of the RA block.

2.4. Parallel Reverse Attention branch

In medical diagnosis, clinicians go for a rough estimation of skin melanoma before looking into the tissue-level finer details for proper localization and labeling. Though, it is not easy for a network to learn residual refinement for saliency detection without proper supervision, leading to inaccurate segmentation results. As most of the existing methods heavily rely on image classification networks, fine-tuned for responsiveness to very few discriminatory regions in images, it deviates from the requirement of exploration of pixel-wise prediction of dense regions. We propose a two-stage segmentation method using a parallel RA unit to mitigate this problem and replicate the real-world clinical approach. The deep layers of the CNN produce coarse-level and a rough estimation of the melanoma region, with small structural details [13]. Next, followed by the idea of progressive erasing of the foreground region [49], we mine discriminative melanoma regions using the RA unit. Instead of aggregating features from all the CNN layers, [16], our proposed RA model guides the learning of the whole

network, starting from the coarse saliency map produced by the deepest CNN layer, containing the highest semantic confidence, by sequentially discovering new information about complementary melanoma regions from the side-output of the last three layers only. Figure 4 shows the architecture of the RA module used in the proposed MFSNet.

Let us consider the last two layers of the CNN have features $F_i; i = \{4, 5\}$, the BA module has output O_B, and the RA mask of the i^{th} level is $\mathcal{M}^i_{RA}$, then the RA output O_R of the i^{th} level is given by Equation 1,where $\mathcal{D}$ is the downsampling operation, $\oplus$ is the concatenation operation of downsampled O_B with the i^{th} level feature set F_i, Con is the operation of passing the feature through a couple of convolutional layers with filter size set to 64, and $\odot$ is the element-wise multiplication of the concatenated feature with the RA mask $\mathcal{M}_{RA}$.

$$O_R(F_i) = \mathcal{M}^i_{RA} \odot Con[F_i \oplus \mathcal{D}(O_B)], \tag{1}$$

Chen et al. [16] defined the RA mask as in Equation 2, where ϵ is the operation of forming a 64-channel tensor by repeating the single-channel output, to match the dimension, $\ominus$ is the subtraction operation, $Softmax$ indicates the sigmoid activation, S_{i+1} is the segmentation mask obtained from the $(i+1)^{th}$ layer of the CNN, $\mathcal{U}$ is the upsampling operation.

$$\mathcal{M}^i_{RA} = \epsilon[1 \ominus Softmax\{\mathcal{U}(S_{i+1}(j)\}], \tag{2}$$

2.5. Boundary Attention

Edge information can guide the task of feature extraction for segmentation by providing helpful supervision with fine-grained boundary constraints as shown in [56]. Hence, being inspired by the Edge Guidance Module (EGM), proposed by [55], we have used a BA module along with the parallel RA branches for extracting accurate boundary information. Based on the fact that only low-level features contain substantial edge information, we have fed the shallow feature F_2 from the encoder network to the BA module as shown in Figure 5.

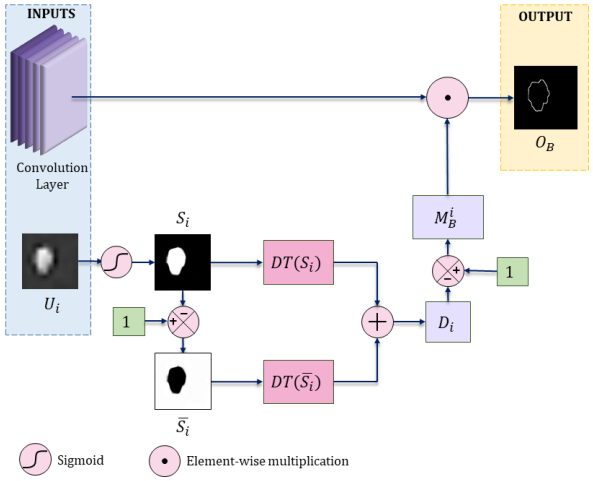


Figure 5: Architecture of the BA module used in the MFSNet model.
U_i: Global map output from PPD module; S_i: Segmentation Map; $\bar{S}_i$: Inverted Segmentation Map; $DT(x)$: Distance Transform; $\mathcal{M}_B^i$: i^{th} level boundary mask; O_B: Boundary Attention output

The BA module helps the network capture important boundary information, which is complementary to the amplified reverse class response for the regions of shared semantic information extracted by the RA module. This additional edge information acts as a helpful signal to confusing segment regions near the lesion boundaries. The i^{th} level feature F_i from the encoder, when fed to the BA module, produces an output O_B, given by Equation 3, where $\odot$ is the element-wise multiplication of feature F_i and the i^{th} level boundary mask $\mathcal{M}_B^i$, which is obtained by formulating the binary segmentation map S_i given by Equation 4, where j is the pixel position index, U_i denotes the i^{th} level upsampled prediction.

$$O_B(F_i) = \mathcal{M}_B^i \odot F_i, \tag{3}$$

$$S_i(j) = \begin{cases} 1, & \text{if } \sigma[U_i(j)] > 0.5 \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

The value of i is set to 2, 3, i.e., we only consider the second and third level

features from the CNN to feed into the BA module. σ is the softmax activation function given by the Equation 5.

$$\sigma(x_i) = \frac{\exp x_m}{\sum_n \exp x_n} \tag{5}$$

Next, distance transformation [18] is applied over S_i to fill each pixel position of the melanoma region with the distance to the melanoma boundary. Conversely, the distances of the pixels of non-melanoma regions can be obtained by simply transposing S_i followed by distance transformation. The overall distance map is produced by normalizing and summing up these two distance maps as given by Equation 6, where $\overline{S_i}$ is the transpose of segmentation map S_i which can be obtained as $\overline{S_i} = 1 - S_i$.

$$D_i = \frac{DT(S_i)}{\max_j DT[S_i(j)]} + \frac{DT(\overline{S_i})}{\max_j DT[\overline{S_i(j)}]}, \tag{6}$$

In Equation 6, D_i has values equal to 0 and 1 at the melanoma boundary and the farthest point from the boundary, respectively. Here, we define the i^{th} level boundary mask ${\mathcal{M}_{\mathcal{B}}}^i$ as

$${\mathcal{M}_{\mathcal{B}}}^i = 1 - D_i \tag{7}$$

Finally, we calculate the boundary map G_B from the ground truth using its gradient, which is constrained by the BCE loss to measure the dissimilarity between the produced boundary map O_B with the actual boundary map G_B given by Equation 8.

$$\mathcal{L}_{\mathcal{B}} = -\sum_j [G_B \log(O_B) + (1 - G_B) \log(1 - O_B)] \tag{8}$$

The overall architecture of the BA module is shown in Figure 5.

2.6. Partial Parallel Decoder module

As suggested by [52], low-level features contribute very little toward the final prediction map with a massive requirement of computation due to their

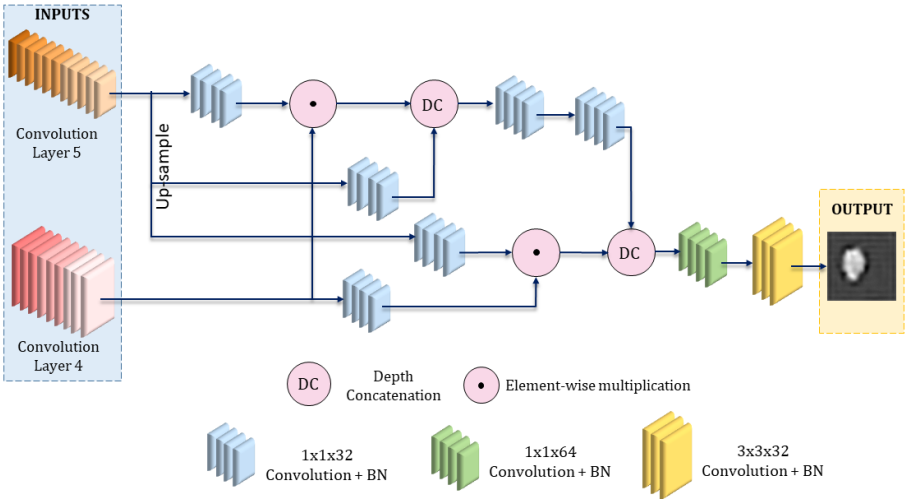


Figure 6: Architecture of the Partial Parallel Decoder module used in the proposed framework. The convolution layers 4 and 5 denote the 4th and fifth layer respectively of the Res2Net CNN backbone used in MFSNet.

high spatial resolution. However, in literature, most of the existing models like [57, 23] are designed to aggregate both high and low-level semantics, leading to unnecessary wastage of resources and inefficient segmentation map. To mitigate this problem, we have used a PPD module to capture the global context information, being inspired from the Receptive Field Block (RFB) module by [29].

Specifically, we have used the first five convolution layers of the Res2Net [20], among which the first three layers are considered as the low-level features and are discarded for the decoder module. To accelerate the feature propagation, we add a series of convolution and batch normalization operations as shown in Figure 6. Short connections are added in the PPD module, similar to the original RFB module. After obtaining different discriminating features from different layers, we finally multiply them to reduce the gap between multiple feature levels. Thus, the PPD module produces a global segmentation map O_S through a series of element-wise multiplication and concatenation operations, serving as the global guidance of the parallel RA branches. Proper downsampling and upsampling

operations are performed throughout, whenever required, to match the feature dimensions before concatenation. Finally, the generated segmentation map is of a similar dimension as that of the input of the MFSNet.

2.7. Deep Supervision

To supervise the segmentation performance, we have used a hybrid loss function in this research. For the BA module, we have used the standard BCE loss function, shown in Equation 8. However, for the supervision of segmentation, we have used a mixing loss function for effective global and local supervision to enhance both image-level and pixel-level segmentation, respectively. The proposed loss function involves the weighted BCE loss function $\mathcal{L}_{wBCE}$ and weighted IoU loss function $\mathcal{L}_{w}IoU$, given by Equation 9, where δ is the weight, set to 0.9 in our case experimentally.

$$\mathcal{L}_S = \delta\mathcal{L}_{wBCE} + (1-\delta)\mathcal{L}_{wIoU}, \tag{9}$$

The experimental analysis is shown in Figure 7. The $\mathcal{L}_{wIoU}$ and $\mathcal{L}_{wBCE}$ are effective to increase the weights of the hard pixels rather than giving equal weights to each pixel like the standard IoU loss and BCE loss functions.

The side outputs from the CNN are upsampled to form segmentation map O_i^{UP}; $i = 4, 5$ of the same size of the ground truth G. Thus, the overall loss function is extended to Equation 10.

$$\mathcal{L} = \mathcal{L}_S(G, O_S) + \sum_{i=2,3} \mathcal{L}_B(G_B, O_{B,i}) + \sum_{i=4,5} \mathcal{L}_S(G, O_i^{UP}) \tag{10}$$

3. Results and Discussion

This section evaluates the proposed framework on three publicly available datasets of skin lesion segmentation, using 5-fold cross-validation. We discuss the significance of the obtained results and compare the model with other state-of-the-art models to justify the superiority of the proposed model.

Figure 7: Experimental analysis of mean DSC and mean IoU on ISIC2017 dataset against different δ values, that defined the weights of different components in the proposed loss function (Equation 9).

3.1. Dataset Description

Three dermatology datasets have been used in the current research to evaluate the performance of MFSNet:

1. PH^2 dataset by [33] consisting of 200 images.
2. ISIC 2017 dataset by [17] consisting of 2379 images.
3. HAM10000 dataset by [42] consisting of 10015 images.

3.2. Evaluation Metrics

To evaluate the performance of the proposed model on the supervised skin lesion segmentation problem, we use five popularly used metrics which are described as follows:

1. Dice Similarity Coefficient (DSC): It is a spatial overlap metric which is computed as in Equation 11 for predicted image S and ground truth G.

$$DSC(S,G) = \frac{2 \times |S \cap G|}{|S| + |G|} \tag{11}$$

2. Intersection over Union (IoU): IoU, also known as Jaccard Index (JI), measures segmentation accuracy by computing the ratio of the intersection of objects and their union when projected on the same plane. Mathematically it is expressed as in Equation 12, where S is the predicted segmentation mask, and G is the original ground truth mask of the image.

$$IoU(S, G) = \frac{|S \cap G|}{|S \cup G|} \tag{12}$$

3. F-Measure (FM): F-Measure is a standard metric that evaluates the harmonic mean of the pixel-wise precision and recall and is mathematically expressed as in Equation 13.

$$FM = \frac{2 \times Precision \times Recall}{Recall + Precision} \tag{13}$$

4. Sensitivity (Sen): It characterizes the percentage of pixels of the object that are accurately classified as the object class and it is computed by Equation 14.

$$Sen(S, G) = \frac{|S \cap G|}{|G|} \tag{14}$$

5. Specificity (Spe): It characterizes the percentage of pixels of the background class that are accurately classified as the background, and it is computed using Equation 15.

$$Spe(S, G) = \frac{|(1 - S) \cap (1 - G)|}{|1 - G|} \tag{15}$$

For all the mentioned evaluation metrics, the mean value over all the test images has been reported in this study for evaluation denoted by $mIoU$, $mDSC$, mFM, $mSen$ and $mSpe$.

3.3. Implementation

The proposed MFSNet is implemented in PyTorch and is accelerated using an NVIDIA Tesla K80 GPU. Table 1 shows the results obtained by MFSNet on the three publicly available datasets using 5-fold cross-validation. The high values of DSC and IoU suggest that the segmentation is reasonably accurate. In contrast, the high Sensitivity and Specificity values suggest the maintenance

Table 1: Results obtained by MFSNet on the three datasets using 5-fold cross-validation.

Dataset	Fold	mDSC	mIoU	mFM	mSen	mSpe
PH^2	1	0.955	0.917	0.947	1.000	1.000
	2	0.956	0.918	0.941	0.991	0.986
	3	0.951	0.915	0.945	0.995	0.999
	4	0.949	0.920	0.943	1.000	1.000
	5	0.958	0.899	0.941	0.989	0.999
	Average	**0.954±0.003**	**0.914±0.008**	**0.944±0.002**	**0.995±0.004**	**0.997±0.002**
ISIC 2017	1	0.991	0.976	0.989	1.000	1.000
	2	0.985	0.971	0.980	1.000	1.000
	3	0.983	0.967	0.980	0.998	0.999
	4	0.986	0.980	0.991	0.999	0.999
	5	0.990	0.975	0.989	0.997	0.998
	Average	**0.987±0.003**	**0.974±0.004**	**0.986±0.005**	**0.999±0.001**	**0.999±0.001**
HAM10000	1	0.911	0.910	0.905	1.000	0.999
	2	0.900	0.901	0.899	0.997	0.998
	3	0.905	0.903	0.906	0.999	1.000
	4	0.904	0.894	0.892	1.000	1.000
	5	0.910	0.900	0.914	0.998	0.998
	Average	**0.906±0.004**	**0.902±0.005**	**0.903±0.007**	**0.999±0.001**	**0.999±0.001**

Table 2: Comparison of the results obtained with the MFSNet model on the three datasets with and without image preprocessing

Dataset	Preprocessing	mDSC	mIoU	mSen	mSpe
PH2	NO	0.931	0.895	0.978	0.978
	YES	**0.954**	**0.914**	**0.995**	**0.997**
ISIC2017	NO	0.963	0.942	0.969	0.970
	YES	**0.987**	**0.974**	**0.999**	**0.999**
HAM10000	NO	0.872	0.869	0.954	0.961
	YES	**0.906**	**0.902**	**0.999**	**0.999**

of structural coherence between the segmented mask and the available ground-truth mask. Further, to evaluate the importance of image preprocessing (artifact removal) in this research, we evaluate and compare the performance of the MFSNet model with the raw images and the preprocessed images. The results of these experiments are presented in Table 2.

3.4. Ablation Study

We have experimented by removing different components from the proposed model to justify their impact on the overall performance. We have performed an ablation study of RA, PPD, BA modules and their different orientations concerning the convolution layers of the backbone Res2Net model to assert the importance of the proposed configuration used in the MFSNet architecture.

3.4.1. Orientation of BA and RA

We have experimented with different combinations and orientations of BA and RA branches to explore the best possible combinations for boosting performance. Table 3 shows the results on the PH^2 dataset, where we have used RA and BA modules at different levels of feature extraction. Comparing instances 1 and 4 from the table shows that the performance can be boosted if we use BA at the $Conv2$ layer instead of RA. This behavior can be justified because the shallow layers of the CNN can extract features rich in boundary information. Hence adding BA there will provide additional edge guidance to the model. Similar conclusions can be drawn by comparing instances 1 and 2. Again, comparing instances 2 and 3, we can observe experimentally that removing the BA module from the $Conv1$ layer does not decrease the segmentation performance significantly but effectively reduces computation of an additional BA module. We have slightly better performance in instance three than in instance 5, establishing the importance of the RA module at the $Conv4$ layer. Based on these observations, we have finalized the orientations of different RA and BA blocks to optimize the segmentation performance and add their clinical importance.

Table 3: Comparison of quantitative results obtained from different orientations of RA and BA blocks in the proposed MFSNet model on the PH^2 dataset. The highlighted row indicates the orientations and results of our proposed model.

Instance	Combinations					Average Result (on 5 fold)				
	Conv1	Conv2	Conv3	Conv4	Conv5	mDSC	mIoU	mFM	mSen	mSpe
1	BA	BA	BA	RA	RA	0.944±0.002	0.897±0.003	0.928±0.004	0.984±0.008	0.989±0.004
2	BA	BA	RA	RA	RA	0.926±.006	0.872±0.002	0.902±0.006	0.971±0.004	0.969±0.006
3	-	BA	RA	RA	RA	0.930±0.004	0.876±0.003	.911±0.007	0.979±0.005	0.972±0.006
4	BA	RA	BA	RA	RA	0.926±0.004	0.876±0.002	0.916±0.005	0.966±0.004	0.963±0.002
5	-	BA	RA	BA	RA	0.926±0.006	0.871±0.005	0.902±0.003	0.978±0.006	0.970±0.004
Proposed	**-**	**BA**	**BA**	**RA**	**RA**	**0.954±0.003**	**0.914±0.008**	**0.944±0.002**	**0.995±0.004**	**0.997±0.002**

Table 4: Results of the ablation study considering various components of the MFSNet model on the PH^2 dataset. Best results are highlighted.

Architecture	mDSC	mIoU	mFM	mSen	mSpe
Res2Net	0.794±0.006	0.758±0.008	0.761±0.005	0.816±0.009	0.821±0.008
Res2Net+PPD	0.877±0.005	0.852±0.004	0.873±0.003	0.915±0.006	0.906±0.004
Res2Net+BA	0.843±0.003	0.820±0.006	0.871±0.004	0.911±0.007	0.904±0.004
Res2Net+RA	0.842±0.002	0.834±0.005	0.866±0.006	0.909±0.006	0.929±0.004
Res2Net+BA+RA	0.906±0.003	0.861±0.004	0.894±0.006	0.947±0.004	0.936±0.007
Res2Net+RA+PPD	0.927±0.003	0.895±0.007	0.912±0.005	0.963±0.007	0.959±0.006
Res2Net+BA+RA+PPD (Proposed)	**0.954±0.003**	**0.914±0.008**	**0.944±0.002**	**0.995±0.004**	**0.997±0.002**

3.4.2. Importance of BA

In this work, we have also performed an ablation study to investigate the importance of the proposed BA module in the overall model. Row 3 in Table 4 shows the performance of the proposed architecture has improved by a considerable margin in terms of significant evaluation metrics by using the BA module along with the *Res2Net* backbone as compared to the backbone only in row 1. Besides, using BA along with RA boosts the model performance as compared to only the RA module, shown in row 4 and row 5 of Table 4, leading to the conclusion that BA has an essential contribution towards achieving better segmentation outcome. [55] also exemplified that optimal edge guidance can boost the segmentation performance significantly, justifying the results obtained in our experiment.

3.4.3. Importance of RA

Row 4 in Table 4 shows that RA is another essential component of the proposed module as removing it may reduce the DSC, IoU, and other evaluation results significantly. The optimal combination of BA and RA modules (shown in Table 3) is another essential feature of the proposed MFSNet, where the addition of RA has boosted the model performance as compared to the mere BA module, shown in row 5 of Table 4.

3.4.4. Importance of PPD

PPD is another vital component of our proposed method, as removal of this can affect the model performance as shown in Table 4. We can observe from row 2 of Table 4 that adding PPD to the baseline model can increase the performance, unparalleled to the contribution of RA and BA. Again, combining it with the RA module, as shown in row 6, can produce an almost similar performance to that of the proposed architecture. The improvements establish that PPD, combined with RA, is the prime component of the proposed MFSNet.

3.5. Comparison to State-of-the-art

Table 5 compares the proposed method to several state-of-the-art methods on the three datasets used. The proposed MFSNet performs significantly better than the said methods and can be justified as a reliable framework for skin lesion segmentation. To further prove the superiority of the MFSNet framework, we use some popular segmentation models prevalent in literature for comparison: U-Net [36], SegNet [5] and Double U-Net [26], the results of which are also compared in Table 5. Some visual results of the predicted segmented masks by these models and the proposed MFSNet are shown in Figure 8. From the visual results, it can be seen that SegNet consistently produces unsatisfactory results. U-Net can segment most images well, but it fails to perform well for relatively challenging images. Double U-Net performs closest to the MFSNet. However evidently MFSNet outperforms all these models as justified from both Table 5 and Figure 8.

Table 5: Comparison of the proposed MFSNet model to state-of-the-art models on the three publicly available datasets used in this study. (Total training time is calculated on implementation using NVIDIA Tesla K80 GPU)

Dataset	Model	mDSC	mIoU	mSen	mSpe	Training time
PH^2	Double U-Net [26]	0.907	0.899	0.945	0.966	1hr 2min12sec
	U-Net [36]	0.876	0.780	0.816	0.978	30min 54 sec
	SegNet [5]	0.894	0.808	0.865	0.966	58min 21 sec
	Goyal et al. [22]	0.907	0.839	0.932	0.929	-
	Hasan et al. [24]	-	0.870	0..929	0.969	35min 08sec
	Al et al. [2]	0.918	0.848	0.937	0.957	-
	Ozturk et al. [35]	0.930	0.871	0.969	0.953	-
	Xie et al. [53]	0.919	0.857	0.963	0.942	-
	Unver et al. [44]	0.881	0.795	0.836	0.940	-
	Yuan et al. [54]	0.915	-	-	-	-
	Bi et al. [11]	0.907	0.840	0.949	0.940	-
	Bi et al. [10]	0.921	0.859	0.962	0.945	-
	Proposed MFSNet	**0.954**	**0.914**	**0.995**	**0.997**	**46min 37sec**
ISIC2017	Double U-Net [26]	0.913	0.918	0.963	0.974	4hr 18min 07sec
	U-Net [36]	0.778	0.683	0.812	0.805	**3hr 22min 44sec**
	SegNet [5]	0.821	0.696	0.801	0.954	4hr 04min 17sec
	Tschandl et al. [43]	0.853	0.770	-	-	-
	Navarro et al. [34]	0.938	0.846	-	-	-
	Saha et al. [37]	0.855	0.772	0.824	0.981	-
	Goyel et al. [22]	0.793	0.871	0.899	0.950	-
	Hasan et al. [24]	-	0.775	0.875	0.955	3hr 37min 17sec
	Al et al. [2]	0.871	0.771	0.854	0.967	-
	Ozturk et al. [35]	0.886	0.783	0.854	0.981	-
	Xie et al. [53]	0.862	0.783	0.870	0.964	-
	Unver et al. [44]	0.843	0.748	0.908	0.927	-
	Proposed MFSNet	**0.987**	**0.974**	**0.999**	**0.999**	3hr 51min 20sec
HAM10000	Double U-Net [26]	0.843	0.812	0.861	0.845	11hr 21min 53sec
	U-Net [36]	0.781	0.774	0.799	0.802	9hr 05min 31sec
	SegNet [5]	0.816	0.821	0.867	0.854	11hr 04min 10sec
	Saha et al. [37]	0.891	0.819	0.824	0.981	-
	Abraham et al. [1]	0.856	-	-	-	-
	Shahin et al. [39]	0.903	0.837	0.902	0.974	-
	Bissoto et al. [12]	0.873	0.792	0.934	0.936	-
	Ibtehaz et al. [25]	-	0.803	-	-	-
	Proposed MFSNet	**0.906**	**0.902**	**0.999**	**0.999**	**9hr 41min 34sec**

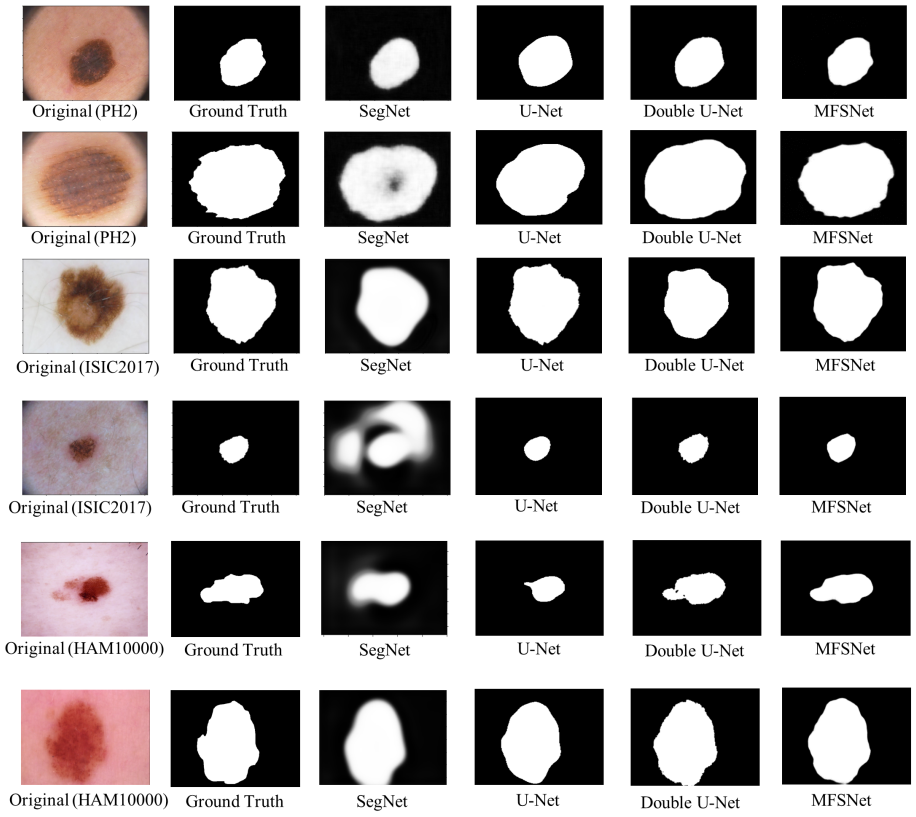


Figure 8: A few instances of segmentation masks obtained by some standard models in literature compared to the proposed MFSNet model.

We have also compared the computational cost of MFSNet in terms of the total execution (training) time with existing methods, as shown in Table 5. It is clear from the table that our proposed method is computationally efficient compared to several state-of-the-art methods like SegNet, DoubleUNet, etc. However, we could not calculate the execution time for all the methods in the literature compared in this study due to the unavailability of open-source implementations.

The improvement in the values of the evaluation metrics by the MFSNet model as compared to state-of-the-art methods in the literature can significantly impact the diagnosis process. Higher values of IoU, DSC, etc., indicate a more accurate skin lesion segmentation while preserving structural similarity. Thus, when the segmented lesions are used for further diagnosis, more robust and informative features can be extracted for the automatic classification of the lesions into benign and malignant classes as stated by Mahbod et al. [32]. This reduces the chances of faulty diagnosis and helps control skin cancer early and more effectively.

4. Conclusion and Future Work

The emergence of CAD systems has facilitated several seemingly daunting tasks, like the segmentation of skin lesions. Skin cancer affects a large population worldwide, and hence its early detection is essential for eradicating cancer. Localization of tumors and lesion segmentation poses a challenge since an esoteric group of clinicians can only perform manual segmentation, and it is also a time-demanding task. To bolster the efforts of the medical practitioners, in this research, we develop a fully automated framework for accurate skin lesion segmentation from raw dermoscopy images. The proposed framework uses multi-scaled maps using a PPD module and two RA and BA modules to produce the final segmentation mask. The use of the multi-focus-based approach helps determine the overall lesion structure from the coarse map. The use of the finer maps helps in determining more refined edges, leading to increased

segmentation accuracy. Upon evaluating the proposed MFSNet model on three publicly available datasets of varied sizes, the proposed method displayed robust performance, outperforming the state-of-the-art methods on the respective datasets.

In the future, we may extend the segmentation model to other domains like brain MRIs, lung CT scans, etc. Also, we might incorporate semi-supervision for the segmentation, similar to [28] to extend the models to unlabelled datasets.

Acknowledgements

The authors would like to thank the Centre for Microprocessor Applications for Training, Education and Research (CMATER) laboratory of the Computer Science and Engineering Department, Jadavpur University, Kolkata, India, for providing the infrastructural support.

Conflict of interest

All the authors declare that there is no conflict of interest.

References

[1] Abraham, N., & Khan, N. M. (2019). A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (pp. 683–687). IEEE.

[2] Al-Masni, M. A., Al-Antari, M. A., Choi, M.-T., Han, S.-M., & Kim, T.-S. (2018). Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Computer methods and programs in biomedicine*, *162*, 221–231.

[3] Aljanabi, M., Özok, Y. E., Rahebi, J., & Abdullah, A. S. (2018). Skin lesion segmentation method for dermoscopy images using artificial bee colony algorithm. *Symmetry*, *10*, 347.

[4] Attia, M., Hossny, M., Nahavandi, S., & Yazdabadi, A. (2017). Skin melanoma segmentation using recurrent and convolutional neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 292–296). IEEE.

[5] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *39*, 2481–2495.

[6] Barata, C., Celebi, M. E., & Marques, J. S. (2021). Explainable skin lesion diagnosis using taxonomies. *Pattern Recognition*, *110*, 107413.

[7] Basak, H., & Kundu, R. (2020). Comparative study of maturation profiles of neural cells in different species with the help of computer vision and deep learning. In *International Symposium on Signal Processing and Intelligent Recognition Systems* (pp. 352–366). Springer.

[8] Beuren, A. T., Janasieivicz, R., Pinheiro, G., Grando, N., & Facon, J. (2012). Skin melanoma segmentation by morphological approach. In *Proceedings of the international conference on advances in computing, communications and informatics* (pp. 972–978).

[9] Bi, L., Feng, D. D., Fulham, M., & Kim, J. (2020). Multi-label classification of multi-modality skin lesion via hyper-connected convolutional neural network. *Pattern Recognition*, *107*, 107502.

[10] Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D., & Fulham, M. (2019). Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern recognition*, *85*, 78–89.

[11] Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., & Feng, D. (2017). Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*, *64*, 2065–2074.

[12] Bissoto, A., Perez, F., Ribeiro, V., Fornaciali, M., Avila, S., & Valle, E. (2018). Deep-learning ensembles for skin-lesion segmentation, anal-

ysis, classification: Recod titans at isic challenge 2018. *arXiv preprint arXiv:1808.08480*, .

[13] Burdick, J., Marques, O., Weinthal, J., & Furht, B. (2018). Rethinking skin lesion segmentation in a convolutional classifier. *Journal of digital imaging*, *31*, 435–440.

[14] Chatterjee, S., Dey, D., & Munshi, S. (2019). Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification. *Computer methods and programs in biomedicine*, *178*, 201–218.

[15] Chattopadhyay, S., & Basak, H. (2020). Multi-scale attention u-net (msaunet): A modified u-net architecture for scene segmentation. *arXiv preprint arXiv:2009.06911*, .

[16] Chen, S., Tan, X., Wang, B., & Hu, X. (2018). Reverse attention for salient object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 234–250).

[17] Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H. et al. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168–172). IEEE.

[18] Fabbri, R., Costa, L. D. F., Torelli, J. C., & Bruno, O. M. (2008). 2d euclidean distance transform algorithms: A comparative survey. *ACM Computing Surveys (CSUR)*, *40*, 1–44.

[19] Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3146–3154).

[20] Gao, S., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., & Torr, P. H. (2019). Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, .

[21] Garcia-Arroyo, J. L., & Garcia-Zapirain, B. (2019). Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding. *Computer methods and programs in biomedicine*, *168*, 11–19.

[22] Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2019). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access*, *8*, 4171–4181.

[23] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., & Liu, J. (2019). Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, *38*, 2281–2292.

[24] Hasan, M. K., Dahal, L., Samarakoon, P. N., Tushar, F. I., & Martí, R. (2020). Dsnet: Automatic dermoscopic skin lesion segmentation. *Computers in Biology and Medicine*, *120*, 103738.

[25] Ibtehaz, N., & Rahman, M. S. (2020). Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, *121*, 74–87.

[26] Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., & Johansen, H. D. (2020). Doubleu-net: A deep convolutional neural network for medical image segmentation. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 558–564). IEEE.

[27] Lei, B., Xia, Z., Jiang, F., Jiang, X., Ge, Z., Xu, Y., Qin, J., Chen, S., Wang, T., & Wang, S. (2020). Skin lesion segmentation via generative adversarial networks with dual discriminators. *Medical Image Analysis*, *64*, 101716.

[28] Li, X., Ma, H., Yi, S., Chen, Y., & Ma, H. (2021). Single annotated pixel based weakly supervised semantic segmentation under driving scenes. *Pattern Recognition*, (p. 107979).

[29] Liu, S., Huang, D. et al. (2018). Receptive field block net for accurate and fast object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 385–400).

[30] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

[31] Ma, Z., & Tavares, J. M. R. (2015). A novel approach to segment skin lesions in dermoscopic images based on a deformable model. *IEEE journal of biomedical and health informatics*, *20*, 615–623.

[32] Mahbod, A., Tschandl, P., Langs, G., Ecker, R., & Ellinger, I. (2020). The effects of skin lesion segmentation on the performance of dermatoscopic image classification. *Computer Methods and Programs in Biomedicine*, *197*, 105725.

[33] Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., & Rozeira, J. (2013). Ph 2-a dermoscopic image database for research and benchmarking. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 5437–5440). IEEE.

[34] Navarro, F., Escudero-Viñolo, M., & Bescós, J. (2018). Accurate segmentation and registration of skin lesion images to evaluate lesion change. *IEEE journal of biomedical and health informatics*, *23*, 501–508.

[35] Öztürk, Ş., & Özkaya, U. (2020). Skin lesion segmentation with improved convolutional neural network. *Journal of digital imaging*, *33*, 958–970.

[36] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference*

on Medical image computing and computer-assisted intervention (pp. 234–241). Springer.

[37] Saha, A., Prasad, P., & Thabit, A. (2020). Leveraging adaptive color augmentation in convolutional neural networks for deep skin lesion segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 2014–2017). IEEE.

[38] Salido, J. A. A., & Ruiz, C. (2018). Using deep learning to detect melanoma in dermoscopy images. *Int. J. Mach. Learn. Comput*, *8*, 61–68.

[39] Shahin, A. H., Amer, K., & Elattar, M. A. (2019). Deep convolutional encoder-decoders with aggregated multi-resolution skip connections for skin lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (pp. 451–454). IEEE.

[40] Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: a cancer journal for clinicians*, *69*, 7–34.

[41] Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of graphics tools*, *9*, 23–34.

[42] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, *5*, 1–9.

[43] Tschandl, P., Sinz, C., & Kittler, H. (2019). Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation. *Computers in biology and medicine*, *104*, 111–116.

[44] Ünver, H. M., & Ayan, E. (2019). Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm. *Diagnostics*, *9*, 72.

[45] Vasconcelos, F. F. X., Medeiros, A. G., Peixoto, S. A., & Reboucas Filho, P. P. (2019). Automatic skin lesions segmentation based on a new morpho-

logical approach via geodesic active contour. *Cognitive Systems Research*, *55*, 44–59.

[46] Verma, K., Singh, B. K., & Thoke, A. (2015). An enhancement in adaptive median filter for edge preservation. *Procedia Computer Science*, *48*, 29–36.

[47] Wang, G., Wang, Y., Li, H., Chen, X., Lu, H., Ma, Y., Peng, C., Wang, Y., & Tang, L. (2014). Morphological background detection and illumination normalization of text image with poor lighting. *PLoS One*, *9*, e110991.

[48] Wang, H., Wang, G., Sheng, Z., & Zhang, S. (2019). Automated segmentation of skin lesion based on pyramid attention network. In *International Workshop on Machine Learning in Medical Imaging* (pp. 435–443). Springer.

[49] Wei, Y., Feng, J., Liang, X., Cheng, M.-M., Zhao, Y., & Yan, S. (2017). Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1568–1576).

[50] Wei, Z., Song, H., Chen, L., Li, Q., & Han, G. (2019). Attention-based denseunet network with adversarial training for skin lesion segmentation. *IEEE Access*, *7*, 136616–136629.

[51] Weng, Y., Zhou, T., Li, Y., & Qiu, X. (2019). Nas-unet: Neural architecture search for medical image segmentation. *IEEE Access*, *7*, 44247–44257.

[52] Wu, Z., Su, L., & Huang, Q. (2019). Cascaded partial decoder for fast and accurate salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3907–3916).

[53] Xie, F., Yang, J., Liu, J., Jiang, Z., Zheng, Y., & Wang, Y. (2020). Skin lesion segmentation using high-resolution convolutional neural network. *Computer methods and programs in biomedicine*, *186*, 105241.

[54] Yuan, Y., Chao, M., & Lo, Y.-C. (2017). Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE transactions on medical imaging*, *36*, 1876–1886.

[55] Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., & Shao, L. (2019). Et-net: A generic edge-attention guidance network for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 442–450). Springer.

[56] Zhao, J.-X., Liu, J.-J., Fan, D.-P., Cao, Y., Yang, J., & Cheng, M.-M. (2019). Egnet: Edge guidance network for salient object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8779–8788).

[57] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer.